Knowledge-Based Systems 72 (2014) 108–113

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A label ranking method based on Gaussian mixture model

Yangming Zhou [a,b], Yangguang Liu [a,*], Xiao-Zhi Gao [c], Guoping Qiu [d]

[a] Ningbo Institute of Technology, Zhejiang University, Ningbo, China
[b] Department of Control Science and Engineering, Zhejiang University, Hangzhou, China
[c] Department of Electrical Engineering and Automation, Aalto University, Finland
[d] School of Computer Science, University of Nottingham, United Kingdom

ARTICLE INFO

Article history:
Received 10 October 2013
Received in revised form 12 July 2014
Accepted 30 August 2014
Available online 16 September 2014

Keywords:
Machine learning
Label ranking
Multi-label learning
Gaussian mixture model
Clustering

ABSTRACT

Label ranking studies the issue of learning a model that maps instances to rankings over a finite set of predefined labels. In order to relieve the cost of memory and time during training and prediction, we propose a novel approach for label ranking problem based on Gaussian mixture model in this paper. The key idea of the approach is to divide the label ranking training data into multiple clusters using clustering algorithm, and each cluster is described by a Gaussian prototype. Then, a Gaussian mixture model is introduced to model the mapping from instances to rankings. Finally, a predicted ranking is obtained with maximum posterior probability. In the experiments, we compare our method with two state-of-the-art label ranking approaches. Experimental results show that our method is fully competitive in terms of predictive accuracy. Moreover, the proposed method also provides a measure of the reliability of the corresponding predicted ranking.

1. Introduction

Label ranking is an increasingly popular topic in the machine learning literature [1–3]. The goal of label ranking is to build a model to describe the mapping from instances to rankings over a finite set of predefined class labels. It extends the conventional classification in the sense that it not only predicts a most likely candidate label but also gives a total order of all class labels. Label ranking problems arise quite naturally in many application areas [4–6]. For example, in movie categorization, where it is likely that a movie belongs to multiple genres (e.g., action, comedy, romance, thriller, etc.), one might not be interested only in predicting which genres are relevant for a specific movie, but also ranking the movie genres by relevance. Other applications include: algorithm recommendation [4], where, given a new data set, the task is to induce a total order of available algorithms according to their suitability based on the data set properties; ranking a set of genes according to their expression level (measured by microarray analysis) based on features of their phylogenetic profiles [7]. Additionally, Fürnkranz et al. [8] proposed a suitable extension of label ranking for solving conventional multi-label classification problems, i.e., introducing a calibration label that represents the boundary between relevant and irrelevant labels.

A number of approaches have been proposed for label ranking learning [2,9–11]. They may be divided into two major categories. One is known as reduction approaches, where a complex label ranking problem is decomposed into several binary classification problems, and then the predictions of these simple models are combined into a ranking of all class labels. For example, Ranking by Pairwise Comparison (RPC), where binary models are learned for each pair of class labels, and the outputs of these models are aggregated into a ranking [12]. The other category is probabilistic approaches, which are based on probabilistic models for ranking, such as Mallows model [13] and Plackett–Luce model [14]. These two models respectively represent two different ways of modeling rankings, i.e., modeling the population of the rankers and modeling the ranking process. A typical representative is Instance-Based (IB) label ranking, including Instance-Based label ranking with Mallows model (IB-M) [15] and the Plackett–Luce model (IB-PL) [16] respectively. Given an instance x, the k-nearest neighbor algorithm is used to find its k neighbors in the instance space. Based on the probabilistic model, all associated rankings of the neighbors are then aggregated to produce the predicted ranking.

Instance-based methods for label ranking fit local model to the data, which are specially useful for problems requiring complex decision boundaries. Additionally, empirical studies [16] show that instance-based label ranking approaches are particularly

* Corresponding author.
E-mail addresses: zhou.yangming@yahoo.com (Y. Zhou), ygliu@acm.org (Y. Liu), xiao-zhi.gao@aalto.fi (X.-Z. Gao), guoping.qiu@nottingham.ac.uk (G. Qiu).

http://dx.doi.org/10.1016/j.knosys.2014.08.029
0950-7051/

competitive to state-of-the-art approaches in terms of predictive accuracy. However, their success is not for free but at a large cost associated with both memory and time. More specifically, both IB-M and IB-PL are lazy learners, with almost no cost at training time but higher cost at prediction time. It can be costly or even impossible in the resources-constrained applications. Moreover, the prediction is based on aggregating the neighbors' rankings. The aggregation is slow as it requires using complex optimization procedures, such as the approximate Expectation Maximization (EM) in IB-M [15] and the Minorization Maximization (MM) in IB-PL [16].

To address this issue, we propose a novel Label Ranking method based on Gaussian Mixture Model (LR-GMM). The key idea of the LR-GMM is to divide the training data into multiple clusters by means of general clustering algorithms, and each cluster is described by a Gaussian prototype. A Gaussian mixture model is then introduced to model the mapping from instances to rankings, and a predicted ranking is obtained with maximum posterior probability. In order to evaluate the performance of LR-GMM, we compare it with two state-of-the-art methods (i.e., RPC and IB-PL) on multiple data sets. Additionally, we also analyze the influence of LR-GMM from three main aspects, including initial centers, standard deviations and the base clustering algorithms. The main contributions made in this work can be summarized as follows:

- Firstly, Gaussian mixture model is used for label ranking learning, and the empirical results are quite promising and suggest that LR-GMM is particularly strong in terms of predictive accuracy.
- Secondly, the modular structure of LR-GMM allows for combining different clustering methods in a convenient way, e.g., different clustering methods can be used according to different application scenarios.
- Thirdly, like other probabilistic model approaches, LR-GMM not only produces a single ranking as an estimation but also delivers a probability distribution over all rankings, which can provide natural measures of the reliability of the predicted rankings.

The paper is organized as follows. In the following section, we briefly describe the label ranking problem in a more formal way. In Section 3, we introduce the proposed label ranking method based on Gaussian mixture model. In Section 4, we provide an experimental study, in which LR-GMM is compared with existing state-of-the-art methods and some analysis of the method has been given. Finally, we close the paper with some concluding remarks in Section 5.

2. Label ranking

The formal description of the label ranking problem given here follows the one provided by Cheng et al. [15]. Label ranking can be considered as a natural extension of conventional classification setting. Instead of associating every instance x from an instance space $\mathcal{X}$ with one among a finite set of class labels $\mathcal{L} = \{\lambda_1, \lambda_2, \ldots, \lambda_m\}$, we associate x with a total order of all class labels. This means that we have a complete, transitive, and asymmetric relation $\succ_x$ on $\mathcal{L}$, where $\lambda_i \succ_x \lambda_j$ indicates that λ_i precedes λ_j in the ranking associated with x. To illustrate, suppose that instances are consumers (characterized by attributes, such as gender, age, occupation and major mobile phones on the market) and $\succ$ is a preference relation on a fixed set of mobile phone brands, such as Apple, Samsung, Nokia, Motorola, and Blackberry.

Formally, a ranking $\succ_x$ can be identified with a permutation π_x of the set $\{1, 2, \ldots, m\}$. It is convenient to define π_x, such that $\pi_x(i) = \pi_x(\lambda_i)$ is the position of λ_i in the ranking, i.e., the rank of λ_i. This permutation encodes the ranking given by

$$\lambda_{\pi_x^{-1}(1)} \succ_x \lambda_{\pi_x^{-1}(2)} \succ_x \cdots \succ_x \lambda_{\pi_x^{-1}(m)} \tag{1}$$

where $\pi_x^{-1}(i)$ is the index of the class label at position i in the ranking. The class of all permutations of the set $\{1, 2, \ldots, m\}$ is denoted by Ω. By abuse of terminology, though justified in light of the above one-to-one correspondence, we refer to elements $\pi \in \Omega$ as both permutations and rankings.

In contrast to classification setting, there is no deterministic mapping from instances to rankings. Instead, every instance is associated with a probability distribution over Ω. That is, for every instance $x \in \mathcal{X}$, there exists a probability distribution $P(\cdot|x)$ such that, for every $\pi \in \Omega, P(\pi|x)$ is the probability with π_x equal to π.

To evaluate the predictive performance of the label ranking algorithm, a proper performance measure is needed. Kendall distance is an important and frequently used measure in label ranking learning [17]. It is based on the number of discordant label pairs,

$$D(\pi, \hat{\pi}) = \#\{(i,j) | \pi(i) > \pi(j) \quad \text{and} \quad \hat{\pi}(i) < \hat{\pi}(j)\} \tag{2}$$

which is closely related to the Kendall's tau correlation (τ). It is a linear scaling of $D(\pi, \hat{\pi})$ to interval $[-1, +1]$, i.e., $\tau = 1 - 4D(\pi, \hat{\pi})/(m(m-1))$. Other alternative performance measures on rankings include Spearman distance and Footrule distance [18]. It has been shown that these three distance measures have tight relations, and thus, any of them could be used [19].

3. Label ranking based on Gaussian mixture model

In this section, we introduce a new label ranking method based on Gaussian mixture model. We assume that the preferences of instances within a cluster are more similar than the preferences of instances in different clusters. It is logical to think that instances sharing the similar characteristics will largely have the similar ranking of all class labels. Based on this assumption, we cluster the label ranking data into K clusters, and each cluster is represented by a Gaussian prototype, i.e., $\{\mu_k, \hat{\pi}_k, \beta_k\}_{k=1,2,\ldots,K}$. More specifically, for the k-th prototype, μ_k is the central position, $\hat{\pi}_k$ is the corresponding center ranking, and β_k is the prior probability of the prototype.

We first introduce the probability $P(k|x)$ of assigning the observation x to the k-th prototype that is dependent on their distance (e.g., Euclidean distance) in the feature space. Let us assume that the probability density $P(x)$ is represented by a mixture model

$$P(x) = \sum_{k=1}^{K} P(x|k) \cdot P(k) \tag{3}$$

where K is the total number of prototypes, $P(k)$ is the prior probability that an instance is generated by a particular prototype, and the Gaussian distribution $P(x|k)$ is the conditional probability that the k-th prototype generates instance x,

$$P(x|k) = \frac{1}{(2\pi\sigma_1^2)^{1/2}} \cdot \exp\left(-\frac{1}{2\sigma_1^2}\|x - \mu_k\|^2\right) \tag{4}$$

Additionally, we assume that all prototypes have the same standard deviation σ_1 in Eq. (4). According to Bayes' rule, we can obtain the assignment probability as follows

$$P(k|x) = \frac{\exp\left(-\|x - \mu_k\|^2/2\sigma_1^2\right) \cdot p(k)}{\sum_{u=1}^{K} \exp\left(-\|x - \mu_u\|^2/2\sigma_1^2\right) \cdot p(u)} \tag{5}$$